Abstract

A fundamental question in a sequential decision making setting under uncertainty is "how to allocate resources amongst competing entities so as to maximize the rewards accumulated in the long run?". The resources allocated may be either abstract quantities such as time or concrete quantities such as manpower. The sequential decision making setting involves one or more agents interacting with an environment to procure rewards at every time instant and the goal is to find an optimal policy for choosing actions. Most of these problems involve multiple (infinite) stages and the objective function is usually a long-run performance objective. The problem is further complicated by the uncertainties in the system, for instance, the stochastic noise and partial observability in a single-agent setting or private information of the agents in a multi-agent setting. The dimensionality of the problem also plays an important role in the solution methodology adopted. Most of the real-world problems involve high-dimensional state and action spaces and an important design aspect of the solution is the choice of knowledge representation.

The aim of this thesis is to answer important resource allocation related questions in different real-world application contexts and in the process contribute novel algorithms to the theory as well. The resource allocation algorithms considered include those from stochastic optimization, stochastic control and reinforcement learning. A number of new algorithms are developed as well. The application contexts selected encompass both single and multi-agent systems, abstract and concrete resources and contain high-dimensional state and control spaces. The empirical results from the various studies performed indicate that the algorithms presented here perform significantly better than those previously proposed in the literature. Further, the algorithms presented here are also shown to theoretically converge, hence guaranteeing optimal performance.

We now briefly describe the various studies conducted here to investigate problems of resource allocation under uncertainties of different kinds:

Vehicular Traffic Control The aim here is to optimize the 'green time' resource of the individual lanes in road networks that maximizes a certain long-term performance objective. We develop several reinforcement learning based algorithms for solving this problem. In the infinite horizon discounted Markov decision process setting, a Q-learning based traffic light control (TLC) algorithm that incorporates feature based representations and function approximation to handle large road networks is proposed, see Prashanth and Bhatnagar [2011b]. This TLC algorithm works with coarse information, obtained via graded thresholds, about the congestion level on the lanes of the road network. However, the graded threshold values used in the above Q-learning based TLC algorithm as well as several other graded threshold-based TLC algorithms that we propose, may not be optimal for all traffic conditions. We therefore also develop a new algorithm based on SPSA to tune the associated thresholds to the 'optimal' values (Prashanth and Bhatnagar [2012]). Our threshold tuning algorithm is online, incremental with proven convergence to the optimal values of thresholds. Further, we also study average cost traffic signal control and develop two novel reinforcement learning based TLC algorithms with function approximation (Prashanth and Bhatnagar [2011c]). Lastly, we also develop a feature adaptation method for 'optimal' feature selection (Bhatnagar et al. [2012a]). This algorithm adapts the features in a way as to converge to an optimal set of features, which can then be used in the algorithm.

Service Systems The aim here is to optimize the 'workforce', the critical resource of any service system. However, adapting the staffing levels to the workloads in such systems is nontrivial as the queue stability and aggregate service level agreement (SLA) constraints have to be complied with. We formulate this problem as a constrained hidden Markov process with a (discrete) worker parameter and propose simultaneous perturbation based simulation optimization algorithms for this purpose. The algorithms include both first order as well as second order methods and incorporate SPSA based gradient estimates in the primal, with dual ascent for the Lagrange multipliers. All the algorithms that we propose are online, incremental and are easy to implement. Further, they involve a certain generalized smooth projection operator, which is essential to project the continuous-valued worker parameter updates obtained from the SASOC algorithms onto the discrete set. We validate our algorithms on five real-life service systems and compare their performance with a state-of-the-art optimization tool-kit OptQuest. Being 25 times faster than OptQuest, our scheme

is particularly suitable for adaptive labor staffing. Also, we observe that it guarantees convergence and finds better solutions than OptQuest in many cases.

Wireless Sensor Networks The aim here is to allocate the 'sleep time' (resource) of the individual sensors in an intrusion detection application such that the energy consumption from the sensors is reduced, while keeping the tracking error to a minimum. We model this sleep–wake scheduling problem as a partially-observed Markov decision process (POMDP) and propose novel RL-based algorithms - with both long-run discounted and average cost objectives - for solving this problem. All our algorithms incorporate function approximation and feature-based representations to handle the curse of dimensionality. Further, the feature selection scheme used in each of the proposed algorithms intelligently manages the energy cost and tracking cost factors, which in turn, assists the search for the optimal sleeping policy. The results from the simulation experiments suggest that our proposed algorithms perform better than a recently proposed algorithm from Fuemmeler and Veeravalli [2008], Fuemmeler et al. [2011].

Mechanism Design The setting here is of multiple self-interested agents with limited capacities, attempting to maximize their individual utilities, which often comes at the expense of the group's utility. The aim of the resource allocator here then is to efficiently allocate the resource (which is being contended for, by the agents) and also maximize the social welfare via the 'right' transfer of payments. In other words, the problem is to find an incentive compatible transfer scheme following a socially efficient allocation. We present two novel mechanisms with progressively realistic assumptions about agent types aimed at economic scenarios where agents have limited capacities. For the simplest case where agent types consist of a unit cost of production and a capacity that does not change with time, we provide an enhancement to the static mechanism of Dash et al. [2007] that effectively deters misreport of the capacity type element by an agent to receive an allocation beyond its capacity, which thereby damages other agents. Our model incorporates an agent's preference to harm other agents through a additive factor in the utility function of an agent and the mechanism we propose achieves strategyproofness by means of a novel penalty scheme. Next, we consider a dynamic setting where agent types evolve and the individual agents here again have a preference to harm others via capacity misreports. We show via a counterexample that the dynamic pivot mechanism of Bergemann and Välimäki [2010] cannot be directly applied in our setting with capacity-limited

agents. We propose an enhancement to the mechanism of Bergemann and Välimäki [2010] that ensures truthtelling w.r.t. capacity type element through a variable penalty scheme (in the spirit of the static mechanism). We show that each of our mechanisms is ex-post incentive compatible, ex-post individually rational, and socially efficient.